Recognizing Emotions and Sentiments in Text: A Survey

Diksha Sahni, Gaurav Aggarwal
Department of CSE
ITM University,
Gurgaon , Haryana, India

Abstract- Emotions are key semantic component of human communication. Due to increasing amount of textual information such as blogs, discussion forums, review sites, chatting data available on the web. Now a days most of the research is focusing on the area of emotion detection and sentiment analysis. Now a days researchers are intended to develop a system that can classify emotion as represented in text. This paper focuses on studying different work done in this area and provides a review that covers the methods and techniques used in emotion detection of text.

I. INTRODUCTION

Emotion is energy-in-motion. It is a way of expressing oneself in life. Emotions are essential part of human interaction. The emotion and sentiment recognition mechanism has been implemented in many kinds of media Some of them are facial expressions, speech ,image ,textual data and so on. Among these textual data is of great importance to the researchers. Due to its small storage medium textual data is the most appropriate medium for network transmission. Now a days most of the communication is performed through text. This kind of communication includes SMS communication, chat communication, emails as well as review or feedback based communication. .A wide range of research work is related to emotions and sentiments in fields like communication, psychology, linguistics. Basically the emotions are divided into two broad types i.e. Positive and Negative. There are six basic categories of emotions namely happiness, sadness, anger, disgust, surprise and fear. Examples of positive emotions include happiness, laughter, interest and so on. Negative emotions include anger ,fear, sadness etc.

Text often reflects the writer's emotional state which can be identified by using various methods .Recognizing emotions and sentiments from text is a classification task which comes under text mining. A large number of statistical and machine learning classifier techniques has been developed for affective computing, which includes decision tree, Naive Bayes, vector-space-model, Support vector machines and so on.
In order to recognize emotions from the given text, the system studies the input text very carefully. It has wide applications in several fields:

- **Opinion analysis**

Opinion analysis aims to receive positive and negative comments about the product as reviews. In this way it helps many companies to examine the opinion of their customers.

- **Better human computer interaction systems**

In human computer interaction systems, emotion recognition techniques are applied to recognize the emotions of a user and to make the system feel like human.

- **Text to speech generation.**

To generate speech from a text, it is needed to analyze the emotion behind the text.in this way the recognition of emotions from text becomes the constructive area of research for speech generation.

Data mining is the process of analyzing correlations or patterns among numerous fields in large databases. It analyze data from different perspectives and summarize it into useful information.

Text Mining is an essential theme in data mining and is all about looking for patterns in a text. These days most of the information related to business, government, industry and other institutions are stored electronically in the form of text databases. Text databases such as news, digital libraries, research papers, articles, e-mail messages, and web pages consists of large portion of available information.

Emotions in text are subjective expressions that describe the feelings of people events, entities and their properties. In most cases, it is difficult for a reader to extract relevant sentences with emotions as emotions are hidden behind the text. Most research in textual emotion sensing target documents which is known as document-level affect detection.

Sentiment analysis is a natural language processing and information extraction task that aims to obtain writer's feelings that are expressed in the form of positive and negative comments. Generally sentiment analysis is used to determine the attitude of writer with respect to some topic. The web is a huge repository of structured and unstructured data. The analyzing this type of data to extract public opinion and sentiment is a challenging task.[1]

A. Data Sources

With an increasing usage of internet, the area of online communication is growing rapidly. People share their opinions, experiences about whatever concerns them online. Sentiment analysis is commonly used by advertisers, movie creators and other organizations that wish to acquire their customer's reaction on a specific topic. Review sites, blogs, micro blogs, data set provides a good understanding for analyzing emotions and sentiments in text.[2]

Review Sites

A review site is a website on which reviews can be posted about various products, people, businesses etc. For a purchasing decision the opinions of customers is considered as an essential factor. A large amount of public generated reviews are available on the web. The public review data in most of the opinion mining study is collected from the e-commerce websites like www.flipkart.com, www.amazon.com, www.reviewcentre.com which hosts millions of product reviews to and also provide ratings for the product.[3]

Blogs

With the widespread use of web ,the trend of writing blog pages is growing rapidly.Now a days writing blog pages have become the most popular means to express one's personal opinions on the web. Bloggers write their opinions, feelings in a blog. Most of these blogs contain reviews of many products and services. In this way blogs are used as data source in most of the studies that are related to sentiment analysis.[3]

Micro-Blogging.

Twitter is a most popular micro-blogging site where people can create their status messages called "tweets". Their tweets express different views and opinions about several topics. These status messages are used as data source for recognizing emotions on the web.

B. Basic Architecture

A lot of work is done has been done in the area of emotion recognition using emotion keyword dictionary based approaches and finding the positive and negative emotion from text. The basic architecture[4] is shown in the figure:
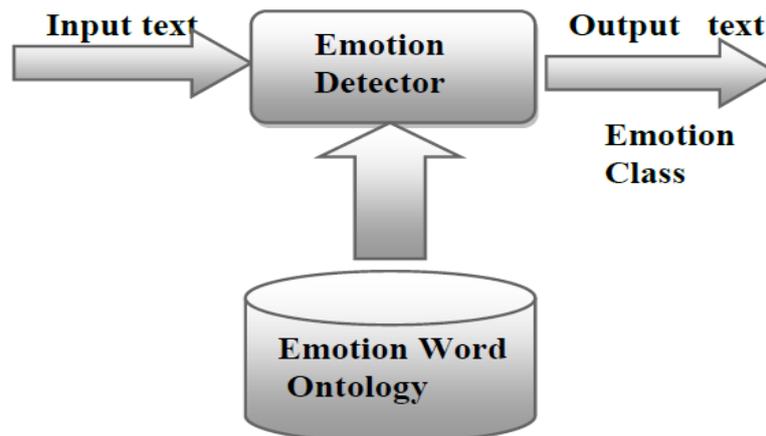

Figure. 1

The basic architecture is divided into two main parts: Emotion word Ontology and Emotion Detector.

An ontology is defined as "Specification of a conceptualization" i.e. ontology describes concepts and relationships that exist for an entity or a community of entities. Ontologies provide an understanding of particular domain. The emotion word ontology is converted into emotion heirarchy. In emotion hierarchy, emotion classes at the primary level are kept on the top of emotion ontology and emotion classes at the tertiary level are kept at the bottom of ontology. Upper level emotion classes are assigned with high weight and lower level emotion classes with low weight.

A emotion detection algorithm is used to recognize the emotions of textual data. The emotion detector algorithm assigns weight for each emotion by adding weights assigned at each level of hierarchy.

C. Computational Approaches of emotion detection in text

There are three approaches for emotion detection task: Keyword based approach,learning Based approach and hybrid Based approach.To detect emotions in these approaches syntactic and semantic data such as n-grams, pos tags etc are used.

Keyword based approach involves identifying words to search for in textual data. It is straight forward and easy to implement.This approach relies on specific emotional keywords and needs pre processing for better accuracy results.

Learning based approach in emotion detection make use of trained classifier. It classify input textual data into emotion classes by using keywords as features. The disadvantage of using this approach is that it results in blurred boundaries between emotion classes and also lacks context analysis. In this approach support vector machine(SVM) is used to clarify the classification task.It has attained good performance in several fields related to text mining.

Hybrid based approach in emotion detection is a combination of keyword and learning based approaches.This approach provides higher accuracy results by training many classifiers and adding knowledge rich linguistic information from word dictionaries[5].

II. RELATED WORK

Yong-Soo Seol[6] proposed a hybrid system that uses two methods, keyword based and machine learning method. keyword based method uses emotional keywords to determine the emotional state in text. If there is no emotional words then the proposed system uses KBANN.KBANN stands for knowledge based artificial neural network which uses approximate domain knowledge and high level features. The system handles eight kinds of emotion(anger, hope, sadness, love, thank, fear, happiness, neutral).In this system thousands of sentences are used for training and evaluation. The accuracy of recognition of emotion in sentences with emotional keywords is about 90 percent for each emotion. The accuracy in sentences without emotional keywords is around 47-65 percent. Since the accuracy of keyword based method is very high as compared to machine learning method, the accuracy of combined system is better than that of using only KBANN.

In [7] the author presented the sentiment analysis for language learning using Naive Bayes Classifier. The system used data that is a random sample of Facebook statuses. The proposed system is designed to classify an opinion using sentence-level classification whether it entails positive, negative or neutral emotions. The system is trained to accept inputs in the form of status updates. In this study three classifiers were used namely Naive Bayes, Rocchio and Perceptron to predict whether the facebook status is positive or negative. The author collected around 7000 status updates from 90 users.The classifiers were compared in terms of three metrics: precision, recall and F-score. Based on F-score ,Rocchio classifier has the best performance and perceptron the least. Naive bayes performed best in all the 3 cases.so accuracy in analyzing the sentimental state of facebook users, using the Naive Bayes classifier is really high.

Tejasvini patil[8] proposed a novel approach for emotion estimation from the text entered by user on social networking sites.The author developed a visual image generation approach that generates images according to emotion in text.The work in this paper deals with finding out the emotion based on affective words in the sentences and for sentences that does not contain any emotional words.The data set used in the proposed system contains blog and twitter data.The author used Naive Bayes classifier for training the input data.The training set consists of 1000 sentences of each emotion category happy,sad,anger and fear.After comparing the three metrics i.e. precision,recall and f-measure the result indicates that the proposed system works better than the traditional methods.

Ricardo A. Calix[9] proposed a system that automatically identify emotional state in text which can be used to render facial expressions.In this the author used a corpus of children's stories.This study used supervised machine learning technique to classify children stories into one of predefined emotion classes.In this study each sentence is represented as a feature vector and classified using Support vector machines(SVM).For a descriptive sentence ,if the emotion class is identified then the system can use that class to adjust the face parameters that can render the face expression which are associated with the emotion class.After that the performance was measured by using recall and precision accuracy scores.

Lily Dey[10] proposed a system of emotion extraction from real time chat messenger. In order to extract emotions from the messenger, the author built a lexicon of emotion conveying words. The proposed system was encoded using nine emotion labels namely anger, fear, disgust, joy, interest, guilt, sadness, surprise and shame. The entire system was implemented in c#. In this paper the author used rule based approach called vector space model(VSM).VSM is mostly used in information retrieval systems. In VSM each document is represented as a vector and each dimension corresponds to a separate term. The result showed by the application are as follows :overall precision 56.37% for simple sentences,42.71% for compound sentences.The author achieved better outcomes in terms of accuracy as compared to other traditional approaches.

Sivaraman Sriram[11] proposed a machine learning approach called Decision tree to classify the given dataset into two known classes of data. In this system the, dataset used is Digg Dataset from cyber emotion and Sem Eval Affective Text are used in a three month complete crawl. These datasets are designed to have six different emotional strength, positive-negative sentimental strength and emotional intensity.The emotional texts are gathered from both of the datasets.In this approach the Root mean square and mean of each dimension in emotion class are evaluated.After implementation the accuracy of the Customized decision tree(CDT) is proved to be a lot better than other existing works.

 Chee kian long[12] explored the potential application of text mining and sentiment analysis for short message service(SMS) texts in teaching evaluation.The system involves three models:the base model,the corrected model which fixes spelling errors and the sentiment model which extends the corrected model by carrying out sentiment mining.

III. EVALUATION AND DISCUSSION

The efficiency of different methods used in recognizing emotions and sentiments in text is measured by three metrics: Precision, Recall and F-score[13].Precision is a measure of accuracy which specifies that a specific class has been predicted. Recall is a measure of capability of the proposed system to select instances of a particular class. F-score Is defined as the harmonic mean of precision and recall measures. The overview of the work done in the emotion detection is given in Table I. This table represents a summary of work done and the work published on the topic of recognizing emotions in text.

TABLE I

Author	Task Description	Emotion Model	DataSet	Accuracy
Yong-Soo Seol	Emotion detection from text using knowledge base artificial neural network	KBANN	Sentences	-
Christos Troussas,Maria Virvou	Sentiment analysis of facebook statuses for language learning.	Naive Bayes Classifier	Facebook statuses	By comparing 3 classifiers naive bayes yield the best accuracy.
Tejasvini patil and Sachin Patil	Automatic generation of emotions for social networking websites using text mining	Naive Bayes Classifier	Blogs,tweets and chatting data	89.38%
Ricardo A.Calix and Bin Chen	Emotion detection in text for 3-D facial expression rendering	SVM	Children stories	70.55%
Lily Dey,Nadia Afroz and Rudra Pratap Deb Nath	Emotion extraction from real time chat messenger	Vector space model	Chatting data	44.70%
Sivaraman Sriram and Xiaobu Yuan	Classifying emotions using customized decision tree algorithm	Decision trees	Digg Dataset	84.36%

IV. RESEARCH METHODOLOGY

At the early stage of work, the domain constructs are identified. The sentiment aspects are identified as the positive and negative adjectives and different weightage is assigned to these adjectives based on the criticality vectors. After assigning these weightage, the next work is to process the single message. This process includes the sentence filtration and identification of domain aspects and the sentiments. Based on the aspect and adjective relationship, the weightage of a particular sentence is identified. At the final stage, classification approach will be applied to identify the feature specific sentiments.

The design of model of the work is represented in following steps:
1) Accept the Domain Specific Dataset
2) Extract the particular Review associated with Domain
3) Define the Stop List of words
4) Identify the keywords present in Message
5) DomainConstructs = Keywords – Positive Adjectives – Negative Adjectives
6) Assign the weightage to adjectives and perform the Domain Aspect Analysis.
7) Identify the weightage of individual message
8) Perform the aggregation on all messages associated with particular Domain
9) Obtain the actual aspect of message in terms of positive or negative sentiment
10) Then the required approach is used to identify the feature specific sentiments

V. CONCLUSION

In this literature survey paper,it is seen that emotion recognition in text plays an important role in human computer interaction systems.Communication in the form of text is the most accustomed approach.The emotion detection from text aims to find the particular emotion by studying the input text.In this paper a survey on various work done in this field is carried out.After survey it is concluded that the Naive bayes classifier is produces the best accuracy.So in future we plan to consider more advanced featureset for emotion classification and recognize emotions more precisely from sentences that contain negation.

REFERENCES

[1]   www.arxiv.org
[2]   G.Vinodhini and RM. Chandrasekaran,"*Sentiment analysis and opinion mining:A survey*" in International journal of advanced research in computer science and software engineering,2012
[3]   http:// www.ijarcsse.com
[4]   Shiv naresh shivhare and prof. Saritha Khethawat ,"*Emotion detection in text*"
[5]   Binali, Haji, Chen Wu, and Vidyasagar Potdar. *"Computational approaches for emotion detection in text"*, 4th IEEE International Conference on Digital Ecosystems and Technologies, 2010.

[6]   Yong-soo seol and Dong-joo kim*,"Emotion recognition from text using Knowledge-based ANN"*,ITC-CSCC 2008

[7]   Christos Troussas,Maria Virvou,"*Sentiment analysis of facebook statuses using Naive Bayes classifier for language learning"*,2012.

[8]   Tejasvini patil and Sachin Patil,"*Automatic generation of emotions for social networking websites using text mining"*in proceedings of IEEE-2012

[9]   Ricardo A.Calix and Bin Chen*,"Emotion recognition in text for 3-D facial expression rendering"*,IEEE-2010

[10]   Lily Dey,Nadia Afroz and Rudra Pratap Deb Nath*,"Emotion extraction from real time chat messenger"*,in proceedings of 3rd International Conference On Informatics,Electronics & Vision 2014.

[11]   Sriram, Sivaraman, and Xiaobu Yuan. "*An enhanced approach for classifying emotions using customized decision tree algorithm*",2012 Proceedings of IEEE Southeastcon,2012.

[12]   Chee Kian Leong and Yew Haur Lee*,"Mining sentiments in SMS texts for teaching evaluation",*Expert systems with applications 39(2012).

[13]   G.Vinodhini and RM. Chandrasekaran," *Sentiment analysis and opinion mining:a survey"*in proceedings of International journal of advanced research in computer science and software engineering,June 2012

[14]   http:// www.wikipedia.com